



Human Leukocyte Antigen Susceptibility Map for Severe Acute Respiratory Syndrome Coronavirus 2

 Austin Nguyen,^{a,b}  Julianne K. David,^{a,b}  Sean K. Maden,^{a,b}  Mary A. Wood,^{a,c}  Benjamin R. Weeder,^{a,b}
 Abhinav Nellore,^{a,b,d}  Reid F. Thompson^{a,b,e,f,g}

^aComputational Biology Program, Oregon Health & Science University, Portland, Oregon, USA

^bDepartment of Biomedical Engineering, Oregon Health & Science University, Portland, Oregon, USA

^cPortland VA Research Foundation, Portland, Oregon, USA

^dDepartment of Surgery, Oregon Health & Science University, Portland, Oregon, USA

^eDepartment of Radiation Medicine, Oregon Health & Science University, Portland, Oregon, USA

^fDepartment of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, USA

^gDivision of Hospital and Specialty Medicine, VA Portland Healthcare System, Portland, Oregon, USA

ABSTRACT Genetic variability across the three major histocompatibility complex (MHC) class I genes (human leukocyte antigen A [HLA-A], -B, and -C genes) may affect susceptibility to and severity of the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for coronavirus disease 2019 (COVID-19). We performed a comprehensive *in silico* analysis of viral peptide-MHC class I binding affinity across 145 HLA-A, -B, and -C genotypes for all SARS-CoV-2 peptides. We further explored the potential for cross-protective immunity conferred by prior exposure to four common human coronaviruses. The SARS-CoV-2 proteome was successfully sampled and was represented by a diversity of HLA alleles. However, we found that HLA-B*46:01 had the fewest predicted binding peptides for SARS-CoV-2, suggesting that individuals with this allele may be particularly vulnerable to COVID-19, as they were previously shown to be for SARS (M. Lin, H.-T. Tseng, J. A. Trejaut, H.-L. Lee, et al., *BMC Med Genet* 4:9, 2003, <https://bmcmmedgenet.biomedcentral.com/articles/10.1186/1471-2350-4-9>). Conversely, we found that HLA-B*15:03 showed the greatest capacity to present highly conserved SARS-CoV-2 peptides that are shared among common human coronaviruses, suggesting that it could enable cross-protective T-cell-based immunity. Finally, we reported global distributions of HLA types with potential epidemiological ramifications in the setting of the current pandemic.

IMPORTANCE Individual genetic variation may help to explain different immune responses to a virus across a population. In particular, understanding how variation in HLA may affect the course of COVID-19 could help identify individuals at higher risk from the disease. HLA typing can be fast and inexpensive. Pairing HLA typing with COVID-19 testing where feasible could improve assessment of severity of viral disease in the population. Following the development of a vaccine against SARS-CoV-2, the virus that causes COVID-19, individuals with high-risk HLA types could be prioritized for vaccination.

KEYWORDS COVID-19, HLA, MHC class I, SARS-CoV-2, coronavirus

Recently, a new strain of betacoronavirus (severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2) emerged as a global pathogen, prompting the World Health Organization in January 2020 to declare an international public health emergency (1). In the large coronavirus family, comprising enveloped positive-strand RNA viruses, SARS-CoV-2 is the seventh encountered strain that causes respiratory disease in

Citation Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF. 2020. Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2. *J Virol* 94:e00510-20. <https://doi.org/10.1128/JVI.00510-20>.

Editor Tom Gallagher, Loyola University Chicago

Copyright © 2020 Nguyen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Abhinav Nellore, nellore@ohsu.edu, or Reid F. Thompson, thompsre@ohsu.edu.

Received 23 March 2020

Accepted 14 April 2020

Accepted manuscript posted online 17 April 2020

Published 16 June 2020

humans (2) ranging from mild—the common cold—to severe—disease caused by the zoonotic Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV). As of April 2020, there are over one million presumed or confirmed cases of coronavirus disease 19 (COVID-19) worldwide, with total deaths exceeding 50,000 (3). While age and many comorbidities, including cardiovascular and pulmonary disease, appear to increase the severity and mortality of COVID-19 (4–9), approximately 80% of infected individuals have mild symptoms (10). As with SARS-CoV (11, 12) and MERS-CoV (13, 14), children seem to have low susceptibility to the disease (15–17); despite infection rates similar to those seen with adults (18), only 5.9% of pediatric cases are severe or critical, possibly due to lower binding ability of the ACE2 receptor in children or generally higher levels of antiviral antibodies (19). Other similarities (20–22), including genomic (23, 24) and immune system response (25–33) similarities, between SARS-CoV-2 and other coronaviruses (34), especially SARS-CoV and MERS-CoV, are topics of ongoing active research, results of which may inform an understanding of the severity of infection (35) and improve the ongoing work of immune landscape profiling (36–40) and vaccine discovery (28, 37, 41–48).

Human leukocyte antigen (HLA) alleles, which are critical components of the viral antigen presentation pathway, have been shown in previous studies to confer differential viral susceptibility and severity of disease. For instance, disease caused by the closely related SARS-CoV (23, 24) shows increased severity among individuals with the HLA-B*46:01 genotype (49). Associations between HLA genotype and disease severity extend broadly to several other unrelated viruses. For example, in human immunodeficiency virus 1 (HIV-1), certain HLA types (e.g., HLA-A*02:05) may reduce risk of seroconversion (50), and in dengue virus, certain HLA alleles (e.g., HLA-A*02:07 and HLA-B*51) are associated with increased secondary disease severity among ethnic Thais (51).

While the details of the clinical picture of the COVID-19 pandemic continue to emerge, there remain substantial unanswered questions regarding the role of individual genetic variability in the immune response against SARS-CoV-2 (51). We hypothesize that individual HLA genotypes may differentially induce the T-cell mediated antiviral response and could potentially alter the course of disease and its transmission. In this study, we performed a comprehensive *in silico* analysis of viral peptide-major histocompatibility complex (MHC) class I binding affinity across 145 different HLA types for the entire SARS-CoV-2 proteome.

RESULTS

To explore the potential for a given HLA allele to produce an antiviral response, we assessed the HLA binding affinity of all possible 8-mers to 12-mers from the SARS-CoV-2 proteome ($n = 48,395$ unique peptides). We then removed from further consideration 16,138 peptides that were not predicted to enter the MHC class I antigen processing pathway via proteasomal cleavage. For the remaining 32,257 peptides, we repeated binding affinity predictions for a total of 145 different HLA types, and we show here the SARS-CoV-2-specific distribution of per-allele proteome presentation (predicted binding affinity threshold of <500 nM) (Fig. 1; see also Table S1 in the supplemental material). Importantly, we note that the putative capacity for SARS-CoV-2 antigen presentation is unrelated to the HLA allelic frequency in the population (Fig. 1). We identify HLA-B*46:01 as the HLA allele with the fewest predicted binding peptides for SARS-CoV-2. We performed the same analyses for the closely related SARS-CoV proteome (see Fig. S1 in the supplemental material) and similarly note that HLA-B*46:01 was predicted to present the fewest SARS-CoV peptides, in keeping with previous clinical data associating this allele with severe disease (49).

To assess the potential for cross-protective immunity conferred by prior exposure to common human coronaviruses (i.e., HKU1, OC43, NL63, and 229E), we next sought to characterize the conservation of the SARS-CoV-2 proteome across diverse coronavirus subgenera to identify highly conserved linear epitopes. After aligning reference proteome sequence data for 5 essential viral components (ORF1ab, S, E, M, and N proteins)

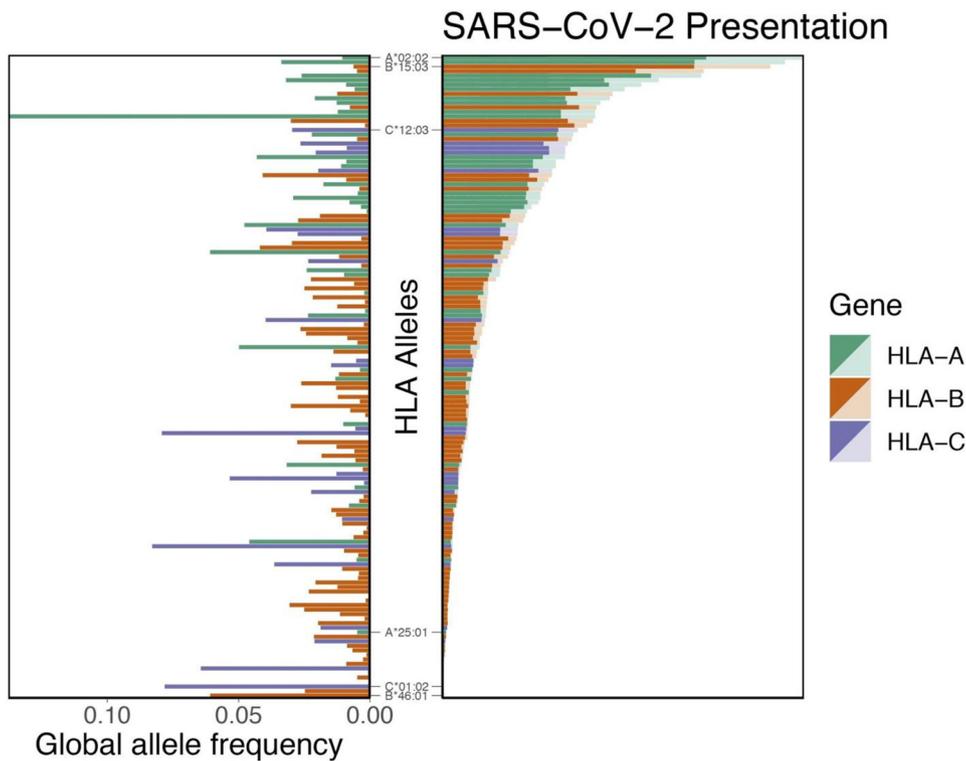


FIG 1 Distribution of HLA allelic presentation of 8- to 12-mers from the SARS-CoV-2 proteome. At right, the number of peptides (see Table S1) that putatively bind to each of 145 HLA alleles is shown as a series of horizontal bars, with dark and light shading indicating the number of tightly (<50 nM) and loosely (<500 nM) binding peptides, respectively, and with green, orange, and purple representing HLA-A, -B, and -C alleles, respectively. Alleles are sorted in descending order based on the number of peptides that they bind (<500 nM). The corresponding estimated allelic frequency in the global population is also shown (left), with the length of each horizontal bar indicating absolute frequency in the population.

across 34 distinct alpha- and betacoronaviruses, including all known human coronaviruses, we identified 48 highly conserved amino acid sequence spans (see Data File S1 in the supplemental material). Acknowledging the challenges to inferring cross-protective immunity among closely related peptides, we confined our attention exclusively to identical peptide matches. Among the conserved sequences, 44 SARS-CoV-2 sequences would each be anticipated to generate at least one 8- to 12-mer linear peptide epitope also present within at least one other common human coronavirus (Fig. 2; see also Table S2). In total, 564 such 8- to 12-mer peptides were found to share 100% identity with corresponding OC43, HKU1, NL63, and 229E sequences (467, 460, 179, and 157 peptides, respectively) (Table S3).

For the subset of these potentially cross-protective peptides that are anticipated to be generated via the MHC class I antigen processing pathway, we performed binding affinity predictions across 145 different HLA-A, -B, and -C alleles (see Data File S3). As described above, we demonstrated the SARS-CoV-2-specific distribution of per-allele presentation for these conserved peptides. We found that alleles HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 were the top presenters of conserved peptides. Conversely, we note that 56 different HLA alleles demonstrated no appreciable binding affinity (<500 nM) to any of the conserved SARS-CoV-2 peptides, suggesting a concomitant lack of potential for cross-protective immunity from other human coronaviruses. We note, in particular, that HLA-B*46:01 was among these alleles. We note also that the putative capacity for conserved peptide presentation is unrelated to the HLA allelic frequency in the population (Fig. 3). Moreover, we see no appreciable global correlation between conservation of the SARS-CoV-2 proteome and its predicted MHC binding affinity, suggesting a lack of selective pressure for or against the capacity to present coronavirus epitopes ($P = 0.27$ [Fisher's exact test]; see Fig. S2).

Nucleocapsid (N) protein

		100	110	120	130	140	150	160	170	180	
<i>SARS-CoV-2</i>	73	PINTNSPDDQIGY	YRRATRR-IRGGD	GKMKDLS	PRWYFY	YLLGTGPE	AGLPYGANKDGI	IIVVATEGALNTP	PKDHIGTRN	PANNAATV	LQLP
<i>SARS-CoV</i>	74	PINTNSGPDDQIGY	YRRATRR-VRGGD	GKMKELS	PRWYFY	YLLGTGPEASL	PYGANKEGI	IIVVATEGALNTP	PKDHIGTRN	PNNNAATV	LQLP
<i>HKU1</i>	86	PIAFGVPPEAKGY	YRHSRRSFKTADG	QQKQLL	PRWYFY	YLLGTGPEANAS	YGESLEGV	FWVANHQADT	STPSDVSS	RDPTTQEA	IPTRFP
<i>OC43</i>	87	PIAPGVPATEAKGY	YRHNRRSFKTADG	NQRQLL	PRWYFY	YLLGTGPHAKDQY	GTIDG	VYVVASNQADV	NTPADIVDR	DPSSDEAI	PTRF
<i>MERS-CoV</i>	64	PLNANSTPAQNA	GWRRODRK-INTG	NG-IKQLA	PRWYFY	YLLGTGPEAAL	PFRAVKDGI	IIVVHEDGAT	DAPSTF	GTRNPN	NSDAIVTQFA
<i>229E</i>	45	PVNKKD-KNKLI	GYWVQKR--FR	TRKGRVDLS	EKLHF	YLLGTGPHKDAK	FRERVE	GVVVAVDG	AKTEPT	-GYGVR	RKNSEPEIPH-F-
<i>NL63</i>	43	PIGKGN-KDEQI	GYWVQER--WR	MRRQRVDL	PKVHF	YLLGTGPHKDL	KFRQRSD	GVVVAKEG	AKTVNT	-SLGNR	KRNQKPLEPK-F-

Membrane (M) protein

		100	110	120	130	140	150	160	170	180	
<i>SARS-CoV-2</i>	85	ACLVGLMWLSYF	IASFRLFARTR	SMWSFN	PETNILL	NVPLHGTIL	TRP	LLESELVIGAVI	LRHLRIAG	HHLGRC-DIKD	LPKEITVATSR
<i>SARS-CoV</i>	84	ACIVGLMWLSYF	VASFRLFARTR	SMWSFN	PETNILL	NVPLRGTIV	TRP	LMESELVIGAVI	IRGLRMAG	HSLGRC-DIKD	LPKEITVATSR
<i>HKU1</i>	86	TIISIVIWILY	FVNSIRLFI	RTGSWWS	FNPE	TNNLMCID	MKGKMFV	RVI	EDYHTLTAT	VIRGHLYIQ	GVKLTGYT
<i>OC43</i>	90	TIVAIMWIVY	FVNSIRLFI	RTGSFWS	FNPE	TNNLMCID	MKGTMYV	REI	EDYHTLT	VTIIRGHLYIQ	GKLTGYSLADL
<i>MERS-CoV</i>	84	AAVSAMMWISY	FVQSIRLFR	MRTGSW	SNPETN	LLNV	PGT	TVVRPL	VEDSTSV	TAVVTN	GHLMKAGMHFGAC-DYDR
<i>229E</i>	84	AVSTLVMWV	MYFANSFRL	FERRART	FWAWN	PEVNAIT	VTTVL	GLQTY	YQIQ	QAPTGIT	VTLLSGVLYVD
<i>NL63</i>	85	SIITLCLW	VMYFVNS	FRLWRR	VKTFW	AFNPETNAI	IISLQ	VYCHNY	YLVMA	APTGVTL	TLSSGVLLVD

ORF1ab polyprotein (Helicase)

		6500	6510	6520	6530	6540	6550	6560	6570	6580	
<i>SARS-CoV-2</i>	5582	ISDEFSSNV	ANYQKVG	MQKYSTL	QPPGTGKSH	FAIGLAL	YPSARIV	YTTACSHA	AVDALCEK	KALKYLP	TDKCSRIIP
<i>SARS-CoV</i>	5559	ISDEFSSNV	ANYQKVG	MQKYSTL	QPPGTGKSH	FAIGLAL	YPSARIV	YTTACSHA	AVDALCEK	KALKYLP	TDKCSRIIP
<i>HKU1</i>	5642	VPLVFQNN	VANYQHIG	MKRYCT	VQPPGTGKSH	LAIGLAV	YTYTAR	VVYTAASHA	AVDALCEK	AYKFLN	INDCTR
<i>OC43</i>	5554	VLETFQNN	VNYQHIG	MKRYCT	VQPPGTGKSH	LAIGLAV	YTYTAR	VVYTAASHA	AVDALCEK	AYKFLN	INDCTR
<i>MERS-CoV</i>	5568	VPEEFASH	VANFQK	SGYSKY	VTVQPPGTGKSH	FAIGLAI	YTYTAR	VVYTTACSHA	AVDALCEK	AFKYLNL	AKCSRIIP
<i>229E</i>	5259	VSDAYANL	VPYQ	LIGKQRI	TIQPPG	SGKSHCS	IGIGV	YYPGAR	IVFTACSHA	AVDSLCAK	AVTAYSVDK
<i>NL63</i>	5229	VSDAYANL	VPYQ	LIGKQRI	TIQPPG	SGKSHCS	IGLGL	YYPGAR	IVFVACA	HAAVDSLCAK	AMTVYS

FIG 2 Amino acid sequence conservation of four linear peptide example sequences from three human coronavirus proteins. Protein sequence alignments are shown for nucleocapsid (N), membrane (M), and ORF1ab polyprotein (helicase) across all five known human betacoronaviruses (*SARS-CoV-2*, *SARS-CoV*, *HKU1*, *OC43*, and *MERS-CoV*) and two known human alphacoronaviruses (*229E* and *NL63*). Each row in the three depicted sequence alignments corresponds to the protein sequence from the indicated coronavirus, with the starting coordinate of the viral protein sequence shown at left and position coordinates of the overall alignment displayed above. Blue shading indicates the extent of sequence identity, with the darkest blue shading indicating a 100% match for that amino acid across all sequences. The four red-highlighted sequences correspond to highly conserved peptides ≥ 8 amino acids in length (*PRWYFYLLGTGPEASL*, *WSFNPETNILL*, *QPPGTGKSH*, and *YTTACSHA*), see Table S2).

We were further interested in whether certain regions of the *SARS-CoV-2* proteome showed differential presentation by the MHC class I pathway. Accordingly, we surveyed the distribution of antigen presentation capacity across the entire proteome, highlighting its most conserved peptide sequences (Fig. 4). Throughout the entire proteome, HLA-A and HLA-C alleles exhibited the relatively largest and smallest capacities to present *SARS-CoV-2* antigens, respectively. However, each of the three major class I genes exhibited very similar patterns of peptide presentation across the proteome (Fig. S3). We additionally note that peptide presentation appears to be independent of estimated time of peptide production during viral life cycle, with indistinguishable levels of peptide presentation of both early and late *SARS-CoV-2* peptides (Fig. S4).

Given the global nature of the current COVID-19 pandemic, we sought to describe population-level distributions of the HLA alleles most (and least) capable of generating a repertoire of *SARS-CoV-2* epitopes in support of a T-cell-based immune response. While we present global maps of individual HLA allele frequencies for the full set of 145 different alleles studied here (Data File S2), we specifically highlight the global distributions of the three best-presenting (*A*02:02*, *B*15:03*, and *C*12:03*) and three of the worst-presenting (*A*25:01*, *B*46:01*, and *C*01:02*) HLA-A, -B, and -C alleles (Fig. 5). Note that all allelic frequencies are aggregated by country but that they implicitly reflect the distribution of HLA data available on the Allele Frequency Net Database (52).

Finally, we acknowledge that nearly all individuals have two HLA-A/B/C haplotypes constituting as few as three but as many as six distinct alleles, potentially buffering against the lack of presentation from a single poorly presenting allele. We sought to determine whether allele-specific variability in *SARS-CoV-2* presentation extends to full

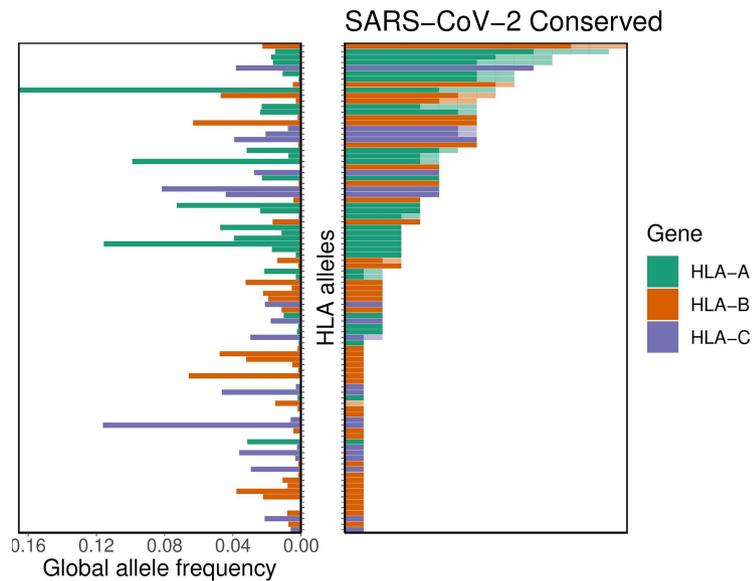


FIG 3 Distribution of HLA allelic presentations of highly conserved human coronavirus peptides with potential to elicit cross-protective immunity to COVID-19. At right, the number of conserved peptides that putatively bind to a subset of 89 HLA alleles is shown as a series of horizontal bars, with dark and light shading indicating the number of tightly (<50 nM) and loosely (<500 nM) binding peptides, respectively, and with green, orange, and purple representing HLA-A, -B, and -C alleles, respectively. Alleles are sorted in descending order based on the number of peptides they are anticipated to present (binding affinity, <500 nM). The corresponding allelic frequency in the global population is also shown (left), with the length of each horizontal bar indicating absolute frequency in the population.

HLA haplotypes and to whole individual HLA genotypes. For six representative alleles with the highest (HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03) and lowest (HLA-A*25:01, HLA-B*46:01, and HLA-C*01:02) predicted capacity for SARS-CoV-2 epitope presentation, these differences remain significant at the haplotype level, albeit with wide variability in presentation among different haplotypes (Fig. 6). Haplotype-level data for all 145 alleles are included in Fig. S5 and Data File S2. We then identified 3,382 individuals with full HLA genotype data and noted wide variability in their capacity to present peptides from the SARS-CoV-2 proteome, albeit with a small minority of individuals at either extreme (Fig. S6).

DISCUSSION

To the best of our knowledge, this was the first study to evaluate per-allele viral proteome presentation across a wide range of HLA alleles using MHC-peptide binding affinity predictors. This report also introduces the relationship between coronavirus sequence conservation and MHC class I antigen presentation. We show that individual HLA, haplotype, and full-genotype variability likely influence the capacity to respond to SARS-CoV-2 infection, and we note certain alleles in particular (e.g., HLA-B*46:01) that could be associated with more-severe infection, as previously shown with SARS-CoV (49). Indeed, we further compare SARS-CoV and SARS-CoV-2 peptide presentation data and note a high degree of similarity between the two across HLA types. Finally, this is the first report to present global distributions of HLA types and haplotypes with potential epidemiological ramifications in the setting of the current pandemic. We found that, in general, there is no correlation between the HLA allelic frequency in the population and allelic capacity to bind SARS-CoV or SARS-CoV-2 peptides, irrespective of the estimated timing of peptide production during the viral replication cycle. While we are not aware of any studies explicitly reporting the relationship between human coronavirus epitope abundance and immune response, there are vaccinia virus data that suggest that early peptide antigens are more likely to generate CD8⁺ T-cell responses whereas antibody and CD4⁺ T-cell responses are more likely to target later mRNA expression with higher peptide abundance in the virion (53).

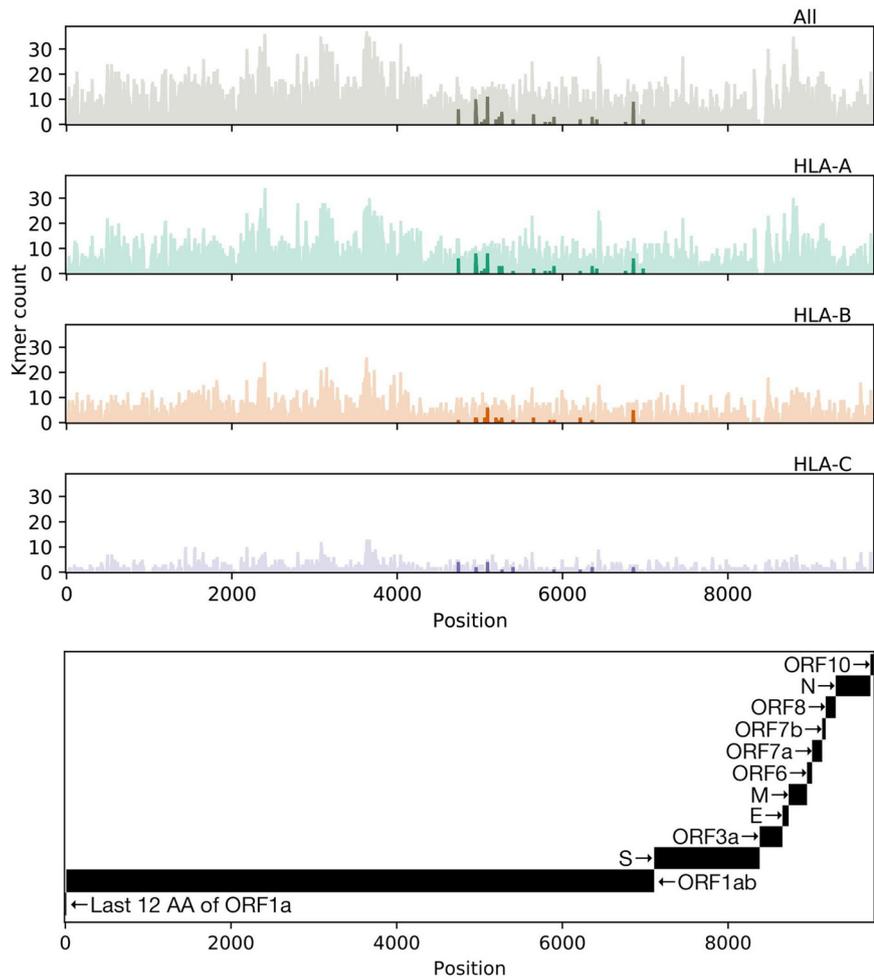


FIG 4 Distribution of allelic presentation of conserved 8- to 12-mers across the entire SARS-CoV-2 proteome for all HLA alleles and individually for HLA-A, HLA-B, and HLA-C (first, second, third, and fourth plots from top, respectively) with dark and light shading indicating the number of tightly (<50 nM) and loosely (<500 nM) binding peptides, respectively. Positions are derived from a concatenation of coding sequences (CDSs) as indicated in the bottom panel. Tightly binding peptides are confined to ORF1ab. The sequence begins with only the last 12 amino acids of ORF1a because all but the last four amino acids of ORF1a are contained in ORF1ab, and we considered binding peptides up to 12 amino acids (AA) in length.

We note, however, several limitations to our work. First and foremost, while we note that a few of our binding affinity predictions were borne out in experimentally validated SARS-CoV peptides (see Table S4 in the supplemental material), we acknowledge that ours was a study performed entirely *in silico*. As we are unable to obtain individual-level HLA typing and clinical outcome data for any real-world COVID-19 populations at this time, the data presented are theoretical in nature and are subject to many of the same limitations implicit in the MHC binding affinity prediction tool(s) upon which it is based. As such, we are unable to assess the relative importance of HLA type compared to known disease-modifying risk factors such as age and clinical comorbidities (4–9). We further note that peptide-MHC binding affinity is limited in its utility as a predictor of subsequent T-cell responses (54–56), and we did not study T-cell responses here. As such, we are ill-equipped to explore phenomena such as original antigenic sin (57–59), where prior exposure to a closely related infection(s) might trigger T-cell anergy (60–62) or immunopathogenesis (63) in the setting of a novel infection. We explored only a limited set of 145 well-studied HLA alleles but note that this analysis could be performed across a wider diversity of genotypes (48). Additionally, we did not assess genotypic heterogeneity or *in vivo* evolution of SARS-CoV-2,

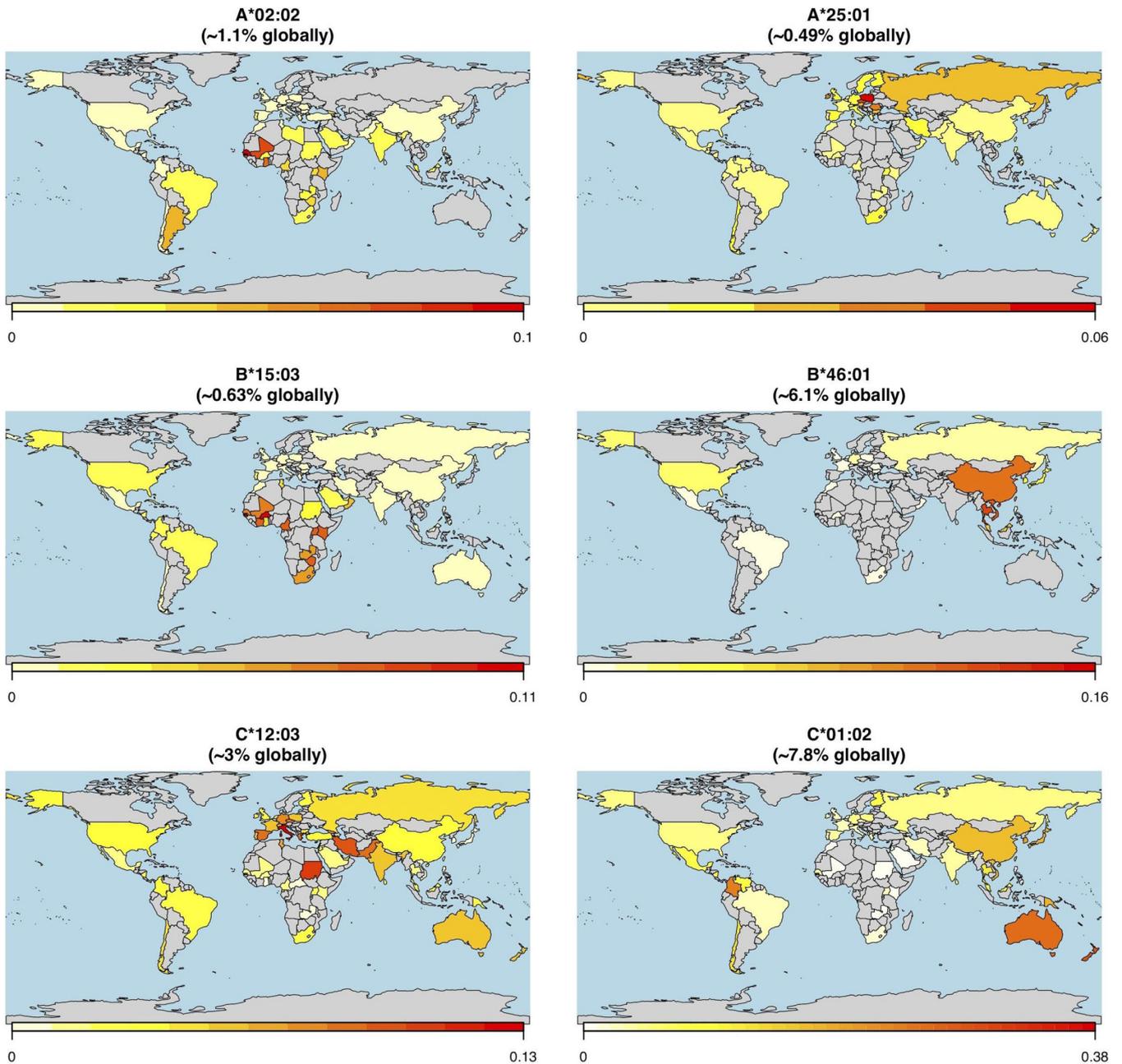


FIG 5 Global HLA allele frequency distribution heat maps for six HLA-A, -B, and -C alleles. The leftmost panels show the global allele frequency distributions by country for three representative alleles (HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03) with the predicted capacities to present the greatest repertoire of epitopes from the SARS-CoV-2 proteome (21.1%, 19.1%, and 7.9% of presentable epitopes, respectively). The rightmost panels show the global allele frequency distributions by country for three representative alleles (HLA-A*25:01, HLA-B*46:01, and HLA-C*01:02) with the lowest predicted levels epitope presentation from the SARS-CoV-2 proteome (0.2%, 0%, and 0% of presentable epitopes, respectively). Heat map coloring corresponds to the individual HLA allele frequency within each country, ranging from lowest (white/yellow) to highest (red) frequency as indicated in the legend below each map.

which could modify the repertoire of viral epitopes presented or could otherwise modulate virulence in an HLA-independent manner (64, 65) (<https://nextstrain.org/ncov>). We also did not address the potential for individual-level genetic variation in other proteins (e.g., angiotensin converting enzyme 2 [ACE2] or transmembrane serine protease 2 [TMPRSS2], essential host proteins for SARS-CoV-2 priming and cell entry [66]) to modulate the host-pathogen interface.

Unless and until the findings we present here are clinically validated, they should not be employed for any clinical purposes. However, we do at this juncture recommend

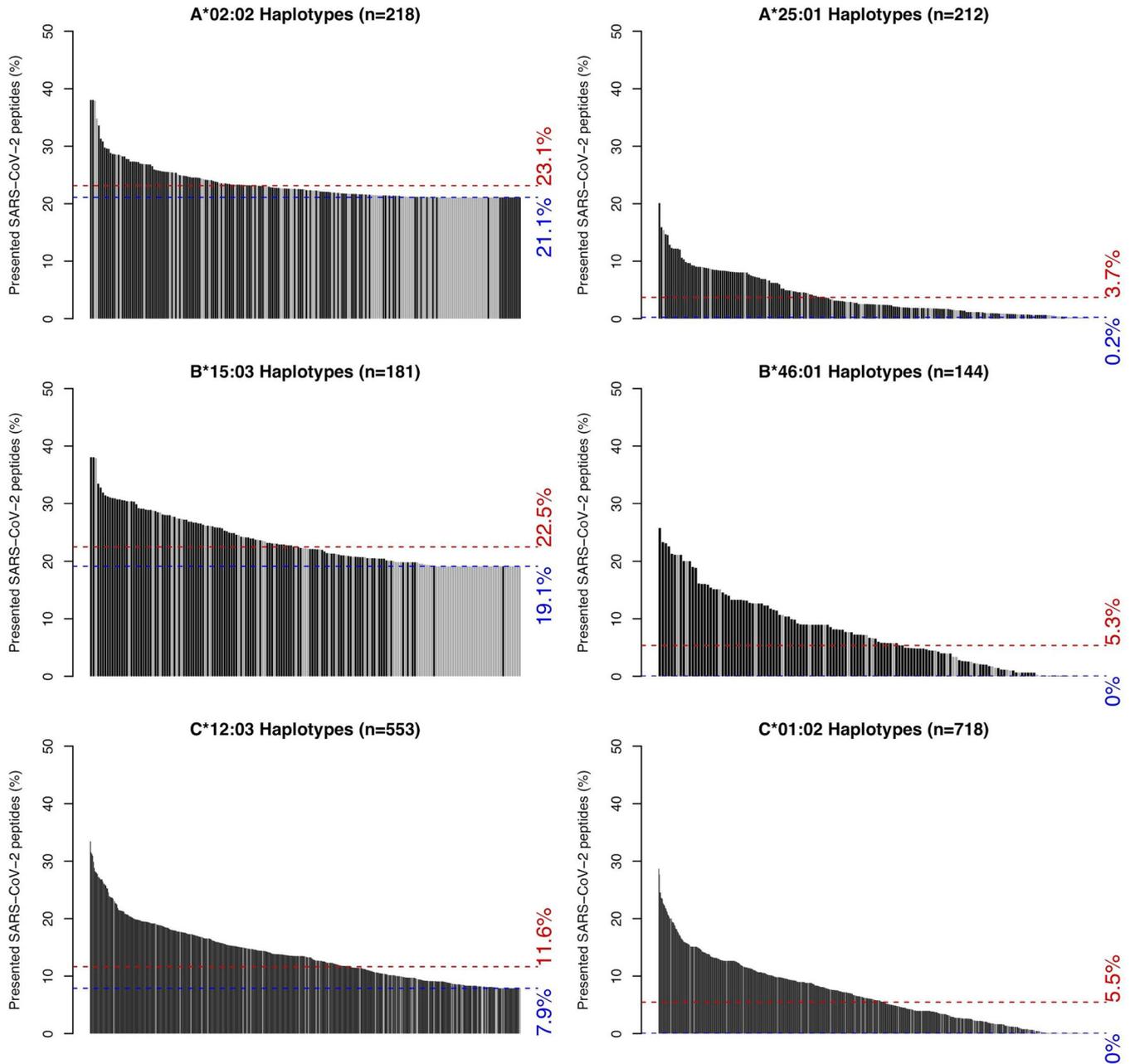


FIG 6 Distributions of SARS-CoV-2 peptide presentation across HLA haplotypes. The leftmost panels show the distributions of SARS-CoV-2 peptide presentation capacity for haplotypes containing one of three representative HLA alleles (HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03) with the greatest predicted repertoire of epitopes from the SARS-CoV-2 proteome. The rightmost panels show the distributions of SARS-CoV-2 peptide presentation capacity for haplotypes containing one of three representative alleles (HLA-A*25:01, HLA-B*46:01, and HLA-C*01:02) with the lowest predicted levels of epitope presentation from the SARS-CoV-2 proteome. Black and gray bars represent full and partial haplotypes, respectively. Blue and red dashed lines represent the percentages of presented SARS-CoV-2 peptides for the indicated allele itself (blue) and its global population frequency weighted average presentation across its observed haplotypes (red).

integrating HLA testing into clinical trials and pairing HLA typing with COVID-19 testing where feasible to more rapidly develop and deploy a predictor(s) of viral severity in the population and, potentially, to tailor future vaccination strategies to genotypically at-risk populations. This approach may have additional implications for the management of a broad array of other viruses.

MATERIALS AND METHODS

Sequence retrieval and alignments. Full polyprotein 1ab (ORF1ab), spike (S) protein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein sequences were obtained for each of 34

distinct but representative alpha and betacoronaviruses from broad genus and subgenus distributions, including all known human coronaviruses (i.e., SARS-CoV, SARS-CoV-2, MERS-CoV, HKU1, OC43, NL63, and 229E). FASTA-formatted protein sequence data (the full accession number list is available in Table S5 in the supplemental material) were retrieved from the National Center of Biotechnology Information (NCBI) (67). For each of the protein classes (i.e., ORF1ab, S, M, E, and N), all 34 coronavirus sequences were aligned using the Clustal Omega v1.2.4 multisequence aligner tool employing the following parameters: sequence type [Protein], output alignment format [clustal_num], dealign [false], mBed-like clustering guide-tree [true], mBed-like clustering iteration [true], number of combined iterations 0, maximum guide tree iterations [-1], and maximum HMM iterations [-1] (68). For the purposes of estimating time of viral peptide production, we classified ORF1a and ORF1b peptides as “early” whereas all other peptides produced by subgenomic mRNAs were classified as “late” (69, 70).

Conserved peptide assessment. Aligned sequences were imported into Jalview v. 2.1.1 (71) with automated generation of the following alignment annotations: (i) sequence consensus, calculated as the percentage of the modal residue per column; (ii) sequence conservation (0 to 11), measured as a numerical index reflecting conservation of amino acid physicochemical properties in the alignment; (iii) alignment quality (0 to 1), measured as a normalized sum of BLOSUM62 ratios for all residues at each position; and (iv) occupancy, calculated as the number of aligned residues (not including gaps) for each position. In all cases, sequence conservation was assessed for each of the following three groups: only human-infecting coronavirus sequences ($n = 7$), all betacoronavirus sequences ($n = 16$), and all alpha- and betacoronavirus sequences combined ($n = 34$). Aligned SARS-CoV-2 sequences and all annotations were manually exported for subsequent analysis. Conserved human coronavirus peptides were defined as those with a length of ≥ 8 consecutive amino acids, each showing agreement with SARS-CoV-2 sequences and ≥ 4 other human coronavirus sequences with the consensus sequence (Table S2). For each of these conserved peptides, we also assessed the component number of 8- to 12-mers sharing identical amino acid sequence between SARS-CoV-2 and each of the four other common human coronaviruses (i.e., OC43, HKU1, NL63, and 229E) (Table S3). For all peptides, human, beta, and combined conservation scores were obtained using a custom R v.3.6.2 script representing mean sequence conservation (minus gap penalties where relevant) (see <https://github.com/pdxgx/covid19>).

Peptide-MHC class I binding affinity predictions. FASTA-formatted input protein sequences from the entire SARS-CoV-2 and SARS-CoV proteomes were obtained from the NCBI RefSeq database (67) under accession numbers [NC_045512.2](https://.ncbi.nlm.nih.gov/nuccore/NC_045512.2) and [NC_004718.3](https://.ncbi.nlm.nih.gov/nuccore/NC_004718.3). We kmerized each of these sequences into 8- to 12-mers to assess MHC class I-peptide binding affinity across the entire proteome. MHC class I binding affinity predictions were performed using 145 different HLA alleles for which global allele frequency data were available as described previously (72) (see Table S5) with netMHCpan v4.0 (73) using the ‘-BA’ option to include binding affinity predictions and the ‘-l’ option to specify peptides 8 to 12 amino acids in length (Table S1). Binding affinity was not predicted for peptides containing the character ‘l’ in their sequences. Additional MHC class I binding affinity predictions were performed on all 66 MHCflurry-supported alleles (–list-supported-alleles; Table S6) using both MHCnuggets 2.3.2 (74) and MHCflurry 1.4.3 (75) (see Tables S7, S8, and S9 and Fig. S7 to S10 in the supplemental material). We further cross-referenced these lists of peptides with existing experimentally validated SARS-CoV epitopes present in the Immune Epitope Database (Table S4) (76). We then performed consensus binding affinity predictions for the 66 supported alleles shared by all three tools by taking the union set of alleles and filtering for peptide-allele pairs matching the union set of alleles. For the SARS-CoV-specific and SARS-CoV-2-specific distributions of per-allele proteome presentation, we exclude all peptide-allele pairs with >500 nM predicted binding. In all cases, we used the netMHCpan v3.0 (77) “C-term” model with a cleavage threshold of 0.1 to further remove any peptides that were not predicted to undergo canonical MHC class I antigen processing via proteasomal cleavage (of the peptide’s C terminus).

Global HLA allele and haplotype frequencies. HLA-A, -B, and -C allele and haplotype frequency data were obtained from the Allele Frequency Net Database (52) for 805 distinct populations pertaining to 101 different countries and 2,628 distinct major/minor (4-digit) alleles, corresponding to 20,478 distinct haplotypes (<https://github.com/pdxgx/covid19>). We also identified full HLA genotype data for 3,382 individuals whose HLA types were confined to the 145 HLA alleles studied here. Population allele and haplotype frequency data were aggregated by country as a mean of all constituent population allele or haplotype frequencies weighted by sample size of the population but not accounting for the representative ethnic demographic size of the population. Global allele frequency maps were generated using the rworldmap v1.3-6 package (78), with total global allele and haplotype frequency estimates calculated as the mean of per-country allele and haplotype frequencies, weighted by each country’s population in 2005.

Data availability. Source code is available at <https://github.com/pdxgx/covid19> under the Massachusetts Institute of Technology (MIT) license. Data File S4 can be found at https://github.com/pdxgx/covid19/blob/master/supporting_data/Appendix_4.zip.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE S1, PDF file, 8.3 MB.

SUPPLEMENTAL FILE S2, PDF file, 9.7 MB.

SUPPLEMENTAL FILE S3, XLS file, 8.8 MB.

SUPPLEMENTAL FILE S4, PDF file, 1.1 MB.

ACKNOWLEDGMENTS

We thank Christopher Loo and Jeffrey Barnet for their critical readings of the manuscript. We thank Jonah Sacha and Paul Spellman for their helpful discussions.

The contents do not necessarily represent the views of the U.S. Department of Veterans Affairs or the United States Government.

R.F.T. was supported by the U.S. Department of Veterans Affairs under award number 1K2CX002049-01 and by the Sunlin & Priscilla Chou Foundation.

REFERENCES

- World Health Organization. 2020. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). WHO (World Health Organization), Geneva, Switzerland.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W; China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- Roser M, Ritchie HR, Ortiz-Ospina E, Hasell J. 2020. Coronavirus disease (COVID-19)—statistics and research. <https://ourworldindata.org/coronavirus>.
- Caramelo F, Ferreira N, Oliveiros B. 2020. Estimation of risk factors for COVID-19 mortality-preliminary results. medRxiv <https://doi.org/10.1101/2020.02.24.20027268>.
- Jain V, Yuan J-M. 2020. Systematic review and meta-analysis of predictive symptoms and comorbidities for severe COVID-19 infection. medRxiv <https://doi.org/10.1101/2020.03.15.20035360>.
- Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y, Pan S, Zou X, Yuan S, Shang Y. 24 February 2020, posting date. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5).
- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323:1061–1069. <https://doi.org/10.1001/jama.2020.1585>.
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395:1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- Guan W-J, Liang W-H, Zhao Y, Liang H-R, Chen Z-S, Li Y-M, Liu X-Q, Chen R-C, Tang C-L, Wang T, et al. 2020. Comorbidity and its impact on 1,590 patients with COVID-19 in China: a nationwide analysis. medRxiv <https://doi.org/10.1101/2020.02.25.20027664>.
- Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. 2020. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* 41:145–151. (In Chinese). <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003>.
- Lau JTF, Lau M, Kim JH, Wong E, Tsui H-Y, Tsang T, Wong TW. 2004. Probable secondary infections in households of SARS patients in Hong Kong. *Emerg Infect Dis* 10:235–243. <https://doi.org/10.3201/eid1002.030626>.
- Denison MR. 2004. Severe acute respiratory syndrome coronavirus pathogenesis, disease and vaccines: an update. *Pediatr Infect Dis J* 23:S207–S214. <https://doi.org/10.1097/01.inf.0000144666.95284.05>.
- Thabet F, Chehab M, Bafaqih H, AlMohaimeed S. 2015. Middle East respiratory syndrome coronavirus in children. *Saudi Med J* 36:484–343. <https://doi.org/10.15537/smj.2015.4.10243>.
- Al-Tawfiq JA, Kattan RF, Memish ZA. 2016. Middle East respiratory syndrome coronavirus disease is rare in children: an update from Saudi Arabia. *World J Clin Pediatr* 5:391–396. <https://doi.org/10.5409/wjcp.v5.i4.391>.
- Cao Q, Chen Y-C, Chen C-L, Chiu C-H. 2020. SARS-CoV-2 infection in children: transmission dynamics and clinical characteristics. *J Formos Med Assoc* 119:670–673. <https://doi.org/10.1016/j.jfma.2020.02.009>.
- Tang A, Xu W, Shen M, Chen P, Li G, Liu Y, Liu L. 2020. A retrospective study of the clinical characteristics of COVID-19 infection in 26 children. medRxiv <https://doi.org/10.1101/2020.03.08.20029710>.
- Lu X, Zhang L, Du H, Zhang J, Li YY, Qu J, Zhang W, Wang Y, Bao S, Li Y, Wu C, Liu H, Liu D, Shao J, Peng X, Yang Y, Liu Z, Xiang Y, Zhang F, Silva RM, Pinkerton KE, Shen K, Xiao H, Xu S, Wong GWK, Chinese Pediatric Novel Coronavirus Study Team. 18 March 2020, posting date. SARS-CoV-2 infection in children. *N Engl J Med* <https://doi.org/10.1056/NEJMco2005073>.
- Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W, Cheng C, Tang X, Wu X, Wu Y, Sun B, Huang S, Sun Y, Zhang J, Ma T, Lessler J, Feng T. 2020. Epidemiology and transmission of COVID-19 in Shenzhen China: analysis of 391 cases and 1,286 of their close contacts. medRxiv <https://doi.org/10.1101/2020.03.03.20028423>.
- Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, Tong S. 16 March 2020, posting date. Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in China. *Pediatrics* <https://doi.org/10.1542/peds.2020-0702>.
- Ashour HM, Elkhatib WF, Rahman MM, Elshabrawy HA. 2020. Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens* 9:E186. <https://doi.org/10.3390/pathogens9030186>.
- Yang Y, Peng F, Wang R, Guan K, Jiang T, Xu G, Sun J, Chang C. 3 March 2020, posting date. The deadly coronaviruses: the 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *J Autoimmun* <https://doi.org/10.1016/j.jaut.2020.102434>.
- Ge Y, Tian T, Huang S, Wan F, Li J, Li S, Yang H, Hong L, Wu N, Yuan E, Cheng L, Lei Y, Shu H, Feng X, Jiang Z, Chi Y, Guo X, Cui L, Xiao L, Li Z, Yang C, Miao Z, Tang H, Chen L, Zeng H, Zhao D, Zhu F, Shen X, Zeng J. 2020. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. bioRxiv <https://doi.org/10.1101/2020.03.11.986836>.
- Ceraolo C, Giorgi FM. 2020. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol* 92:522–528. <https://doi.org/10.1002/jmv.25700>.
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan L, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395:565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- Promptchara E, Ketloy C, Palaga T. 2020. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. *Asian Pac J Allergy Immunol* 38:1–9. <https://doi.org/10.12932/AP-200220-0772>.
- Zheng Z, Monteil VM, Maurer-Stroh S, Yew CW, Leong C, Mohd-Ismael NK, Arularasu SC, Chow VTK, Pin RLT, Mirazimi A, Hong W, Tan Y-J. 2020. Monoclonal antibodies for the S2 subunit of spike of SARS-CoV cross-react with the newly-emerged SARS-CoV-2. bioRxiv <https://doi.org/10.1101/2020.03.06.980037>.
- Yang Y, Shen C, Li J, Yuan J, Yang M, Wang F, Li G, Li Y, Xing L, Peng L, Wei J, Cao M, Zheng H, Wu W, Zou R, Li D, Xu Z, Wang H, Zhang M, Zhang Z, Liu L, Liu Y. 2020. Exuberant elevation of IP-10, MCP-3 and IL-1ra during SARS-CoV-2 infection is associated with disease severity and fatal outcome. medRxiv <https://doi.org/10.1101/2020.03.02.20029975>.
- Ahmed SF, Quadeer AA, McKay MR. 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 12:E254. <https://doi.org/10.3390/v12030254>.
- Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Pan P, Wang W, Hu D, Liu X, Zhang Q, Wu J. 2020. Coronavirus infections and immune responses. *J Med Virol* 92:424–432. <https://doi.org/10.1002/jmv.25685>.

30. Lokugamage KG, Schindewolf C, Menachery VD. 2020. SARS-CoV-2 sensitive to type I interferon pretreatment. *bioRxiv* <https://doi.org/10.1101/2020.03.07.982264>.
31. Wang C, Li W, Drabek D, Okba NMA, van Haperen R, Albert DM, van Kuppeveld FJM, Haagmans BL, Grosveld F, Bosch B-J. 2020. A human monoclonal antibody blocking SARS-CoV-2 infection. *bioRxiv* <https://doi.org/10.1101/2020.03.11.987958>.
32. Lv H, Wu NC, Tsang O-Y, Yuan M, Ranawaka AP, Leung WS, So RTY, Chan JMC, Yip GK, Chik TSH, Wang Y, Choi CYC, Lin Y, Ng WW, Zhao J, Poon LLM, Malik Peiris JS, Wilson IA, Mok CKP. 2020. Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections. *bioRxiv* <https://doi.org/10.1101/2020.03.15.993097>.
33. Yuan M, Wu NC, Zhu X, Lee C-C, So RTY, Lv H, Mok CKP, Wilson IA. 3 April 2020, posting date. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* <https://doi.org/10.1126/science.abb7269>.
34. Tan Y, Schneider T, Leong M, Aravind L, Zhang D. 2020. Novel immunoglobulin domain proteins provide insights into evolution and pathogenesis mechanisms of SARS-related coronaviruses. *bioRxiv* <https://doi.org/10.1101/2020.03.04.977736>.
35. Tetro JA. 2020. Is COVID-19 receiving ADE from other coronaviruses? *Microbes Infect* 22:72–73. <https://doi.org/10.1016/j.micinf.2020.02.006>.
36. Thevarajan I, Nguyen THO, Koutsakos M, Druce J, Cally L, van de Sand CE, Jia X, Nicholson S, Catton M, Iommie B, Tong S, Lewin S, Kedzierska K. 2020. Breadth of concomitant immune responses underpinning viral clearance and patient recovery in a non-severe case of COVID-19. *medRxiv* <https://doi.org/10.1101/2020.02.20.20025841>.
37. Zhu J, Kim J, Xiao X, Wang Y, Luo D, Chen R, Xu L, Zhang H, Xiao G, Zhan X, Wang T, Xie Y. 2020. Profiling the immune vulnerability landscape of the 2019 novel coronavirus. *bioRxiv* <https://doi.org/10.1101/2020.02.08.939553>.
38. Dong C, Ni L, Ye F, Chen M-L, Feng Y, Deng Y-Q, Zhao H, Wei P, Ge J, Li X, Sun L, Wang P, Liang P, Guo H, Wang X, Qin C-F, Chen F. 2020. Characterization of anti-viral immunity in recovered individuals infected by SARS-CoV-2. *medRxiv* <https://doi.org/10.1101/2020.03.17.20036640>.
39. Okba NMA, Muller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, Lamers MM, Sikkema RS, de Bruijn E, Chandler FD, Yazdanpanah Y, Le Hingrat Q, Descamps D, Houhou-Fidouh N, Reusken CBE, Bosch B-J, Drosten C, Koopmans MPG, Haagmans BL. 2020. SARS-CoV-2 specific antibody responses in COVID-19 patients. *medRxiv* <https://doi.org/10.1101/2020.03.18.20038059>.
40. Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, Wang T, Zhang X, Chen H, Yu H, Zhang X, Zhang M, Wu S, Song J, Chen T, Han M, Li S, Luo X, Zhao J, Ning Q. 27 March 2020, posting date. Clinical and immunologic features in severe and moderate coronavirus disease 2019. *J Clin Invest* <https://doi.org/10.1172/JCI137244>.
41. Tahir Ul Qamar M, Rehman A, Ashfaq UA, Awan MQ, Fatima I, Shahid F, Chen L-L. 2020. Designing of a next generation multi-epitope based vaccine (MEV) against SARS-CoV-2: immunoinformatics and in silico approaches. *bioRxiv* <https://doi.org/10.1101/2020.02.28.970343>.
42. Fast E, Chen B. 2020. Potential T-cell and B-cell epitopes of 2019-nCoV. *bioRxiv* <https://doi.org/10.1101/2020.02.19.955484>.
43. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, Makhawi AM. 2020. Design of multi epitope-based peptide vaccine against E protein of human COVID-19: an immunoinformatics approach. *bioRxiv* <https://doi.org/10.1101/2020.02.04.934232>.
44. Baruah V, Bose S. 2020. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J Med Virol* 92:495–500. <https://doi.org/10.1002/jmv.25698>.
45. Robson B. 2020. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. *Comput Biol Med* 119:103670. <https://doi.org/10.1016/j.combiomed.2020.103670>.
46. Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, Chaicumpa W, 18 March 2020, posting date. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Hum Vaccin Immunother* <https://doi.org/10.1080/21645515.2020.1735227>.
47. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. 16 March 2020, posting date. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* <https://doi.org/10.1016/j.chom.2020.03.002>.
48. Campbell KM, Steiner G, Wells DK, Ribas A, Kalbasi A. 2020. Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. *bioRxiv* <https://doi.org/10.1101/2020.03.30.016931>.
49. Lin M, Tseng H-K, Trejaut JA, Lee H-L, Loo J-H, Chu C-C, Chen P-J, Su Y-W, Lim KH, Tsai Z-U, Lin R-Y, Lin R-S, Huang C-H. 2003. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* 4:9. <https://doi.org/10.1186/1471-2350-4-9>.
50. MacDonald KS, Fowke KR, Kimani J, Dunand VA, Nagelkerke NJ, Ball TB, Oyugi J, Njagi E, Gaur LK, Brunham RC, Wade J, Luscher MA, Krausa P, Rowland-Jones S, Ngugi E, Bwayo JJ, Plummer FA. 2000. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 181:1581–1589. <https://doi.org/10.1086/315472>.
51. Stephens HAF, Klaythong R, Sirikong M, Vaughn DW, Green S, Kalayana-rooj S, Endy TP, Libraty DH, Nisalak A, Innis BL, Rothman AL, Ennis FA, Chandanayingyong D. 2002. HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* 60:309–318. <https://doi.org/10.1034/j.1399-0039.2002.600405.x>.
52. González-Galarza FF, Takeshita LYC, Santos EJM, Kempson F, Maia MHT, da Silva ALS, Luiz Teles A, Ghataoraya GS, Alfrévic A, Jones AR, Middleton D. 2015. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res* 43:D784–D788. <https://doi.org/10.1093/nar/gku1166>.
53. Sette A, Grey H, Oseroff C, Peters B, Moutafsi M, Crotty S, Assarsson E, Greenbaum J, Kim Y, Kolla R, Tschärke D, Koelle D, Johnson RP, Blum J, Head S, Sidney J. 2009. Definition of epitopes and antigens recognized by vaccinia specific immune responses: their conservation in variola virus sequences, and use as a model system to study complex pathogens. *Vaccine* 27(Suppl 6):G21–G26. <https://doi.org/10.1016/j.vaccine.2009.10.011>.
54. Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAfee D, Birnbaum ME, Yang X, Su LF, Yu W, Dong S, Gee MH, Jude KM, Davis MM, Groves JT, Goddard WA, Heath JR, Evavold BD, Vale RD, Garcia KC. 2018. Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* 174:672–687.e27. <https://doi.org/10.1016/j.cell.2018.06.017>.
55. Gálvez J, Gálvez JJ, García-Peñarribia P. 2019. Is TCR/pMHC affinity a good estimate of the T-cell response? An answer based on predictions from 12 phenotypic models. *Front Immunol* 10:349. <https://doi.org/10.3389/fimmu.2019.00349>.
56. Zehn D, Lee SY, Bevan MJ. 2009. Complete but curtailed T-cell response to very low-affinity antigen. *Nature* 458:211–214. <https://doi.org/10.1038/nature07657>.
57. Vatti A, Monsalve DM, Pacheco Y, Chang C, Anaya J-M, Gershwin ME. 2017. Original antigenic sin: a comprehensive review. *J Autoimmun* 83:12–21. <https://doi.org/10.1016/j.jaut.2017.04.008>.
58. Park MS, Kim JI, Park S, Lee I, Park M-S. 2016. Original antigenic sin response to RNA viruses and antiviral immunity. *Immune Netw* 16:261–270. <https://doi.org/10.4110/in.2016.16.5.261>.
59. Lambert P-H, Liu M, Siegrist C-A. 2005. Can successful vaccines teach us how to induce efficient protective immune responses? *Nat Med* 11:554–562. <https://doi.org/10.1038/nm1216>.
60. Singh RAK, Rodgers JR, Barry MA. 2002. The role of T cell antagonism and original antigenic sin in genetic immunization. *J Immunol* 169:6779–6786. <https://doi.org/10.4049/jimmunol.169.12.6779>.
61. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, Broadwater A, Kolla RV, De Silva AD, de Silva AM, Mattia KA, Doranz BJ, Grey HM, Shresta S, Peters B, Sette A. 2013. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8 T cells. *Proc Natl Acad Sci U S A* 110:E2046–E2053. <https://doi.org/10.1073/pnas.1305227110>.
62. Klenerman P, Zinkernagel RM. 1998. Original antigenic sin impairs cytotoxic T lymphocyte responses to viruses bearing variant epitopes. *Nature* 394:482–485. <https://doi.org/10.1038/28860>.
63. Rothman AL. 2011. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. *Nat Rev Immunol* 11:532–543. <https://doi.org/10.1038/nri3014>.
64. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, Zhou Z, Yang J, Zhong J, Yang D, Guo L, Zhang G, Li H, Xu Y, Chen M, Gao Z, Wang J, Ren L, Li M. 4 March 2020, posting date. Genomic diversity of SARS-CoV-2 in coronavirus disease 2019 patients. *Clin Infect Dis* <https://doi.org/10.1093/cid/ciaa203>.
65. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J. 3 March 2020, posting date. On the origin and continuing

- evolution of SARS-CoV-2. *Nat Sci Rev* <https://doi.org/10.1093/nsr/nwaa036>.
66. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, Müller MA, Drosten C, Pöhlmann S. 4 March 2020, posting date. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* <https://doi.org/10.1016/j.cell.2020.02.052>.
 67. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
 68. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
 69. Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* 332:498–510. <https://doi.org/10.1016/j.virol.2004.11.038>.
 70. Taiaroa G, Rawlinson D, Featherstone L, Pitt M, Caly L, Druce J, Purcell D, Harty L, Tran T, Roberts J, Scott N, Catton M, Williamson D, Coin L, Duchene S. 2020. Direct RNA sequencing and early evolution of SARS-CoV-2. *bioRxiv* <https://doi.org/10.1101/2020.03.05.976167>.
 71. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
 72. Wood MA, Paralkar M, Paralkar MP, Nguyen A, Struck AJ, Ellrott K, Margolin A, Nellore A, Thompson RF. 2018. Population-level distribution and putative immunogenicity of cancer neoepitopes. *BMC Cancer* 18: 414. <https://doi.org/10.1186/s12885-018-4325-6>.
 73. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. 2017. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 199:3360–3368. <https://doi.org/10.4049/jimmunol.1700893>.
 74. Shao XM, Bhattacharya R, Huang J, Sivakumar IKA, Tokheim C, Zheng L, Hirsch D, Kaminow B, Omdahl A, Bonsack M, Riemer AB, Velculescu VE, Anagnostou V, Pagel KA, Karchin R. 2020. High-throughput prediction of MHC class I and II neoantigens with MHCnuggets. *Cancer Immunol Res* 8:396–408. <https://doi.org/10.1158/2326-6066.CIR-19-0464>.
 75. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. 2018. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst* 7:129–132.e4. <https://doi.org/10.1016/j.cels.2018.05.014>.
 76. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. 2019. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* 47:D339–D343. <https://doi.org/10.1093/nar/gky1006>.
 77. Nielsen M, Lundegaard C, Lund O, Keşmir C. 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57:33–41. <https://doi.org/10.1007/s00251-005-0781-7>.
 78. South A. 2011. rworldmap: a new R package for mapping global data. *R J* 3:35–43. <https://doi.org/10.32614/RJ-2011-006>.